Federated Network for Air Quality Data and Processing Services

Project Coordinators:
Software Architecture: R. Husar
Software Implementation: K. Höijärvi
Data and Applications: S. Falke, R. Husar

Center for Air Pollution Impact and Trend Analysis (CAPITA)
Washington University, St. Louis, MO 63130
June 2005, rhusar@me.wustl.edu

ITR ReaSON

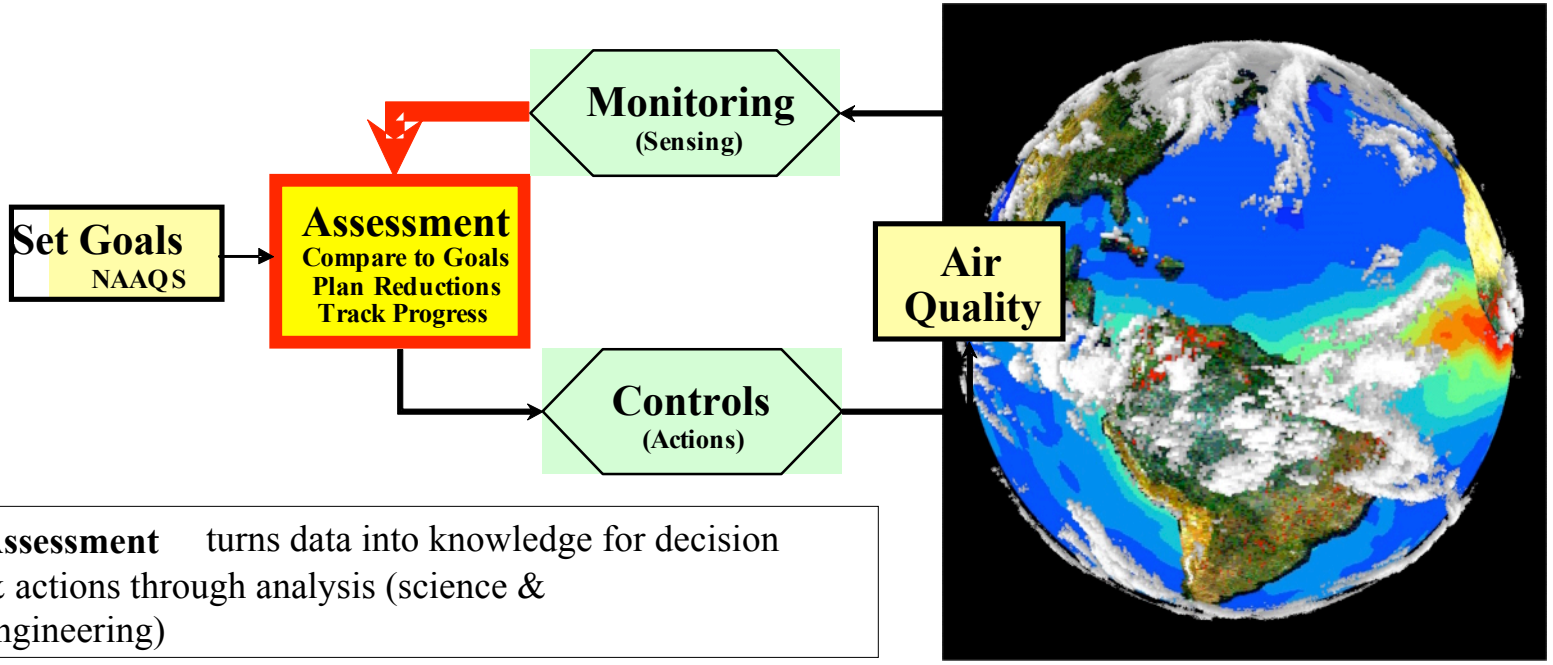
Outline

- Air quality information landscape and processes
- Mediated peer-to-peer DataFed architecture
- Web Services-based distributed applications
- Examples

Air Quality Management Process:

Setting Goals, Monitoring, Assessment, Control

Monitoring collects multi-sensory data from surface and satellite platforms and

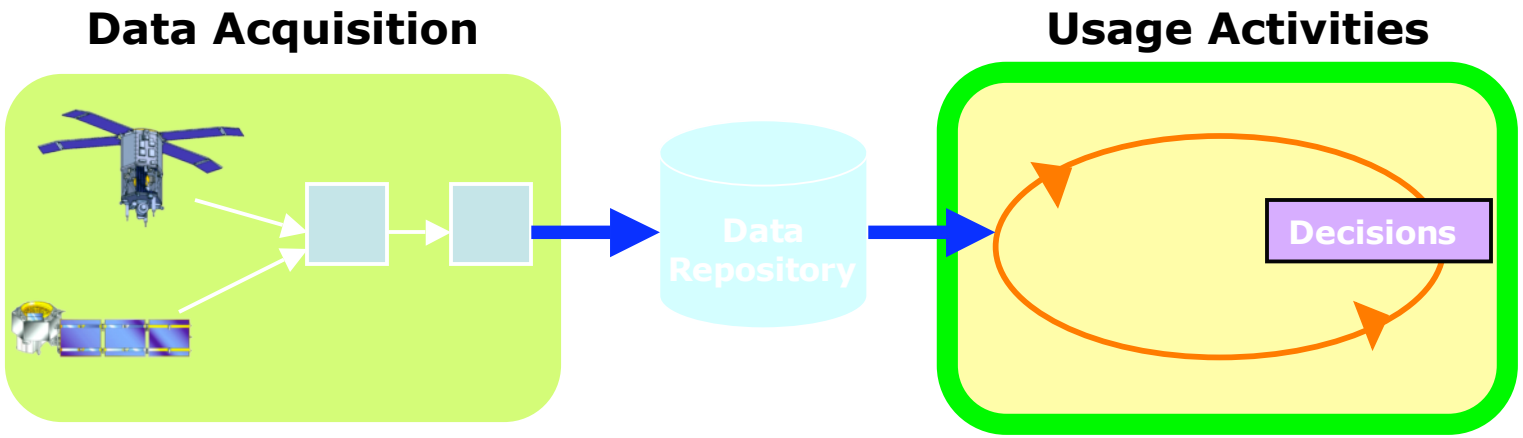


Assessment turns data into knowledge for decision & actions through analysis (science & engineering)

The adoptive AQ management requires and agile supporting info system

ITR ReaSON

Data Acquisition and Usage Activities



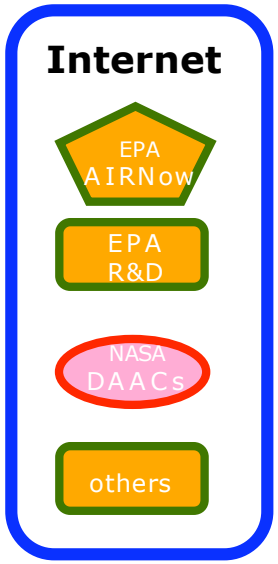
- The data life cycle consists of the **acquisition** and the **usage** parts
- The **acquisition** part processes the sensory data by **firmly** linked procedures
- The collected and cleaned data are **stored** in the repository
- The **usage** activities are more iterative, **dynamic procedures**
- The usage cycle **transform data into knowledge** for decision making

The analysis focuses on data **usage activities** and presumes repositories

ITR ReaSON

Information Landscape: Providers

Geography, Content, Agency, Form

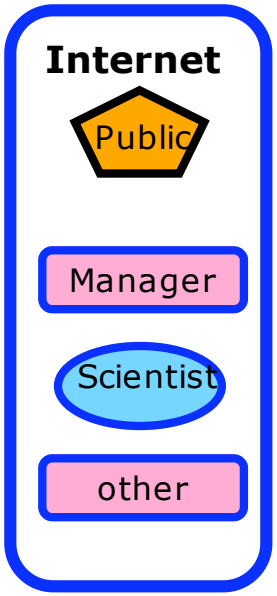


- Data are **distributed** geographically by **autonomous** providers
- Data includes **emissions, ambient** data, **satellite** data and **model** output
- Data are provided by multiple **agencies**: EPA, NOAA, NASA and others
- Furthermore, data are provided in varied **formats** and access **protocols**
- Data on Internet are geography-independent and can be 'linearized'

ITR ReaSON

Information Landscape: Users

Types, Agency, Info Needs



- Users are **distributed** geographically
- Users includes **policy makers,** the **public**, AQ **managers** and **scientist**
- Users are affiliated with multiple **agencies**: EPA, NOAA, NASA, as well as others
- Furthermore, users need various types of information provided in multiple formats
- Users are also on the Internet, their geographic **location is irrelevant**

ITR ReaSON

Information Landscape: Info System

Data Access, Processing and Products

- The info system should transform the data into info products for each user

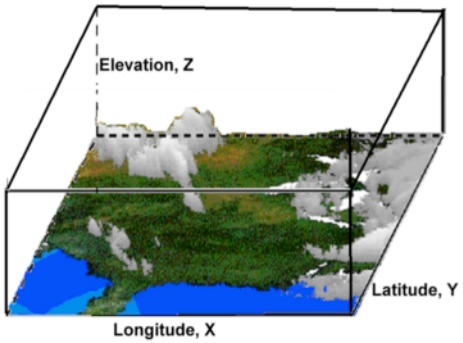
ITR ReaSON

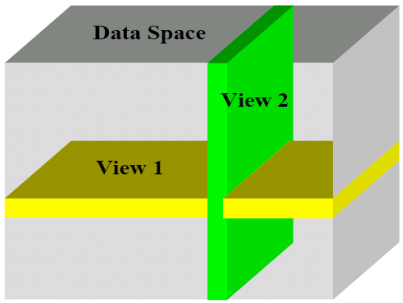
DataFed Multidimensional Data Model

4 D Geo-Environmental Data Cube (X, Y, Z, T)

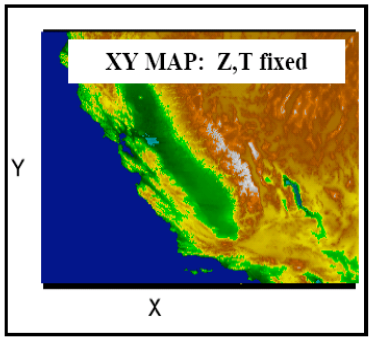
ES data have physical dimension coordinates X, Y, Z, T (**Lat, Lon, Elev., Time)**

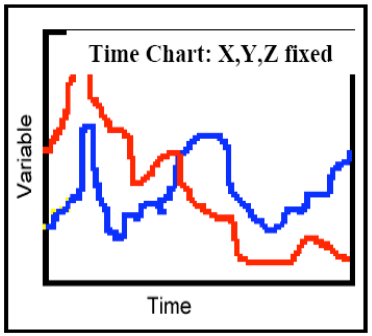
Data exploration and analysis can start by slicing the 4D data cubes

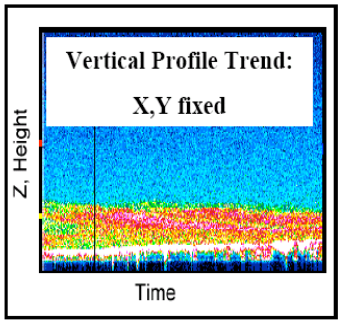


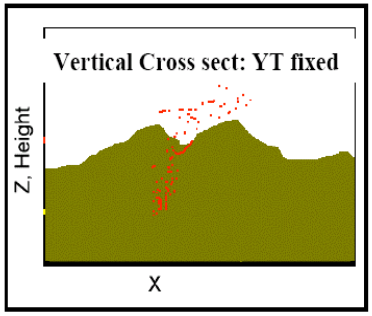


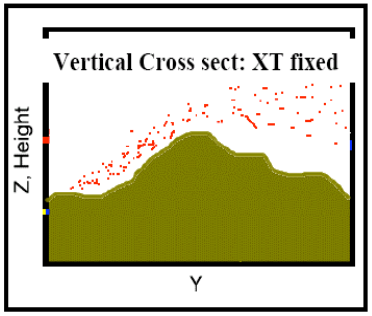
Common Views (slices) through 4D Data Space

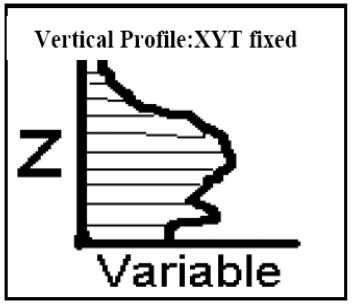








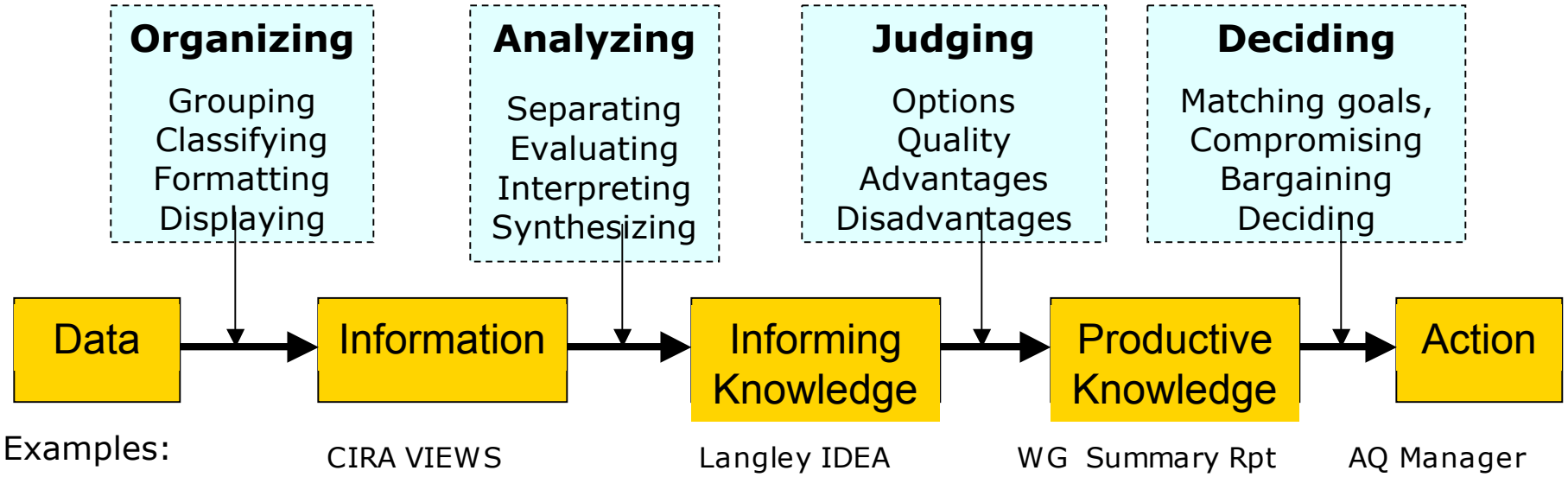




ITR ReaSON

The 'Decision Support System'

(after Taylor, 1987)

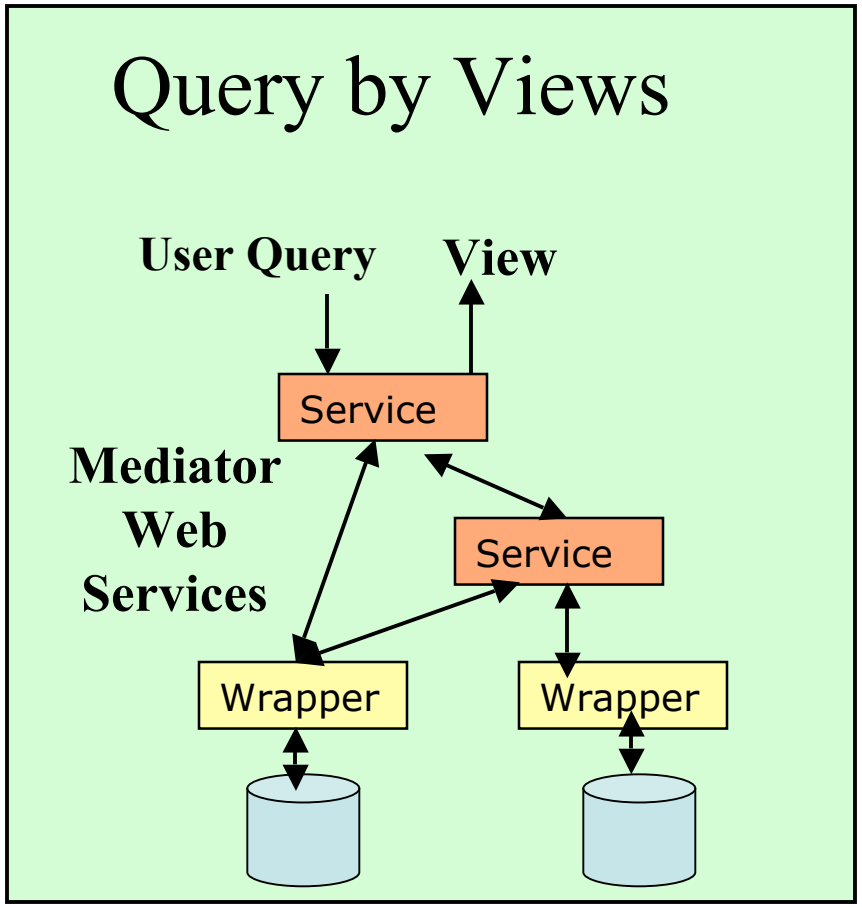


- Processing of AQ data into 'productive' knowledge occurs through a **value-chain**
- Currently, much of the decision support is done by human **analysts and advisers**
- Information technologies could automate the **Data-to-Information transformation**
- This would liberate more resources for **Analyzing and Judging** activities
- More productive or 'actionable' knowledge would lead to **better decision making**

ITR ReaSON

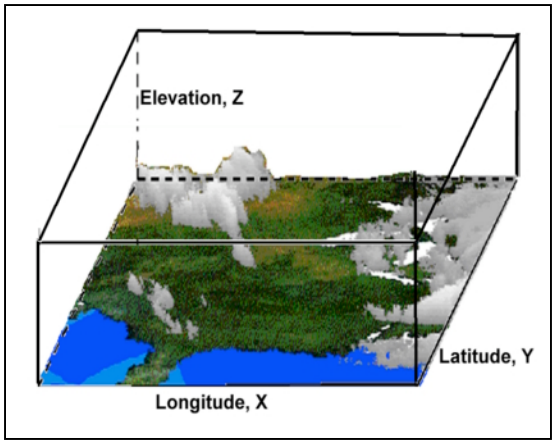
Mediator-Based Integration Architecture
(Wiederhold, 1992)

- Heterogeneous sources are *wrapped* by translating local to **global language**
- *Mediators (web services)* obtain data from wrappers or other mediators
- Mediators provide an answer to a user query in form of *views* (Ullman, 1997)



Data Model:

Geo-Spatio-Temporal Semantics



Layer	Description	Standards
Semantics	Meaning	WSDL ext., Policy, RDF
Data	Types	Schema, WSDL
Protocol	Communication	SOAP, WS-* ext.
Syntax	Data format	XML
Transport	Addressing, Data flow	HTTP, SMTP

Anatomy of a Wrapper Service: TOMS Satellite Image Data

The daily TOMS images reside on the FTP archive, e.g.
ftp://toms.gsfc.nasa.gov/pub/eptoms/images/aerosol/y2000/ea000820.gif

URL template: ftp://toms.gsfc.nasa.gov/pub/eptoms/images/aerosol/y[yyyy]/ea[yy][mm][dd].gif

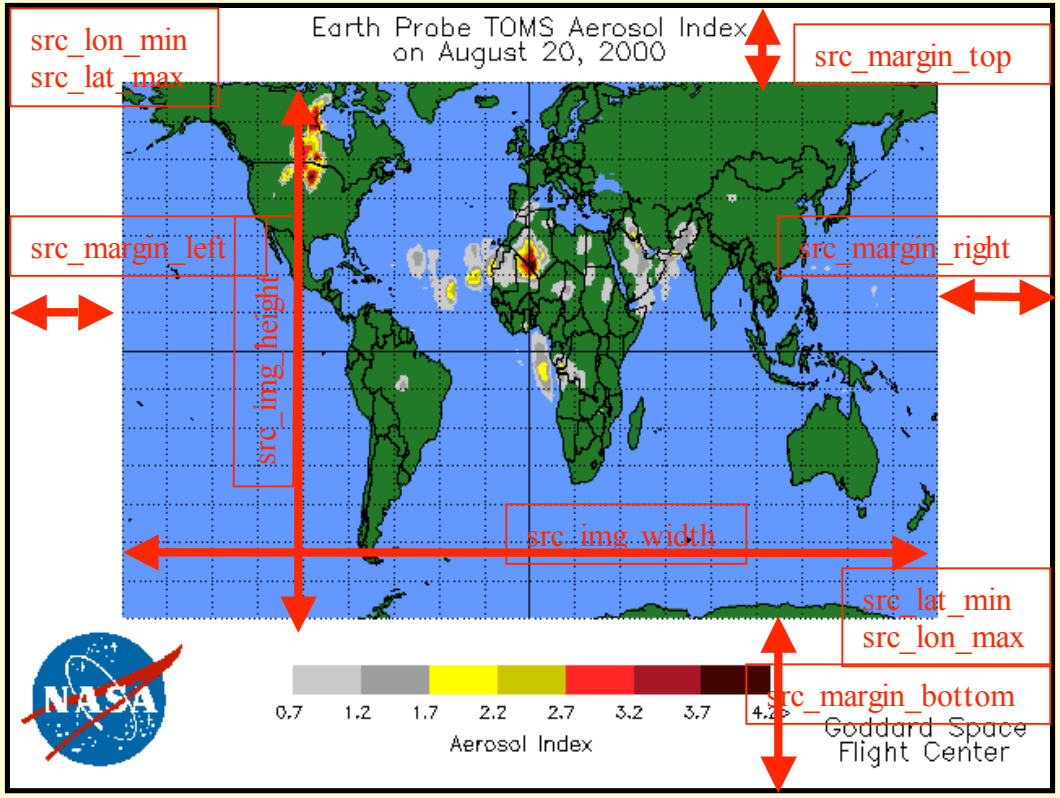


Image Description for Data Access:

```
src_image_width=502
src_image_height=329

src_margin_bottom=105
src_margin_left=69
src_margin_right=69
src_margin_top=46

src_lat_min=-70
src_lat_max=70
src_lon_min=-180
src_lon_max=180
```

Transparent colors for overlays

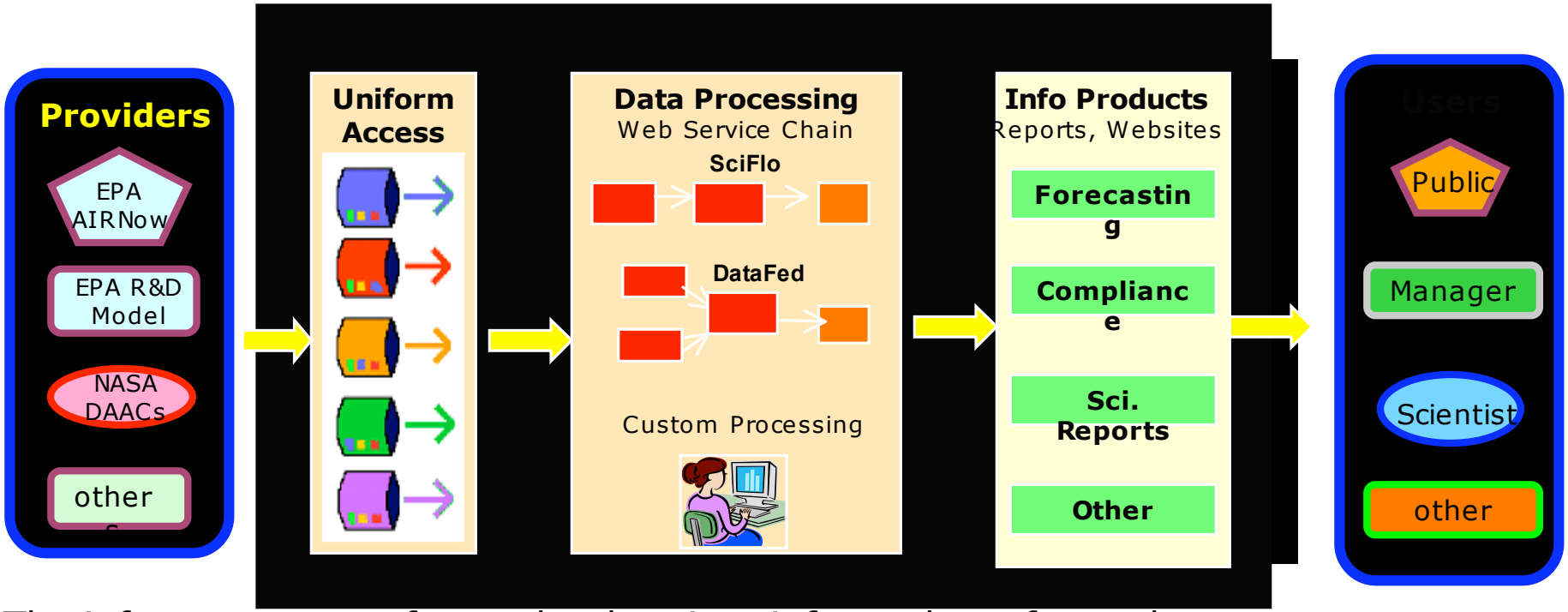
```
RGB(89,140,255)
RGB(41,117,41)
RGB(23,23,23)
RGB(0,0,0)
```

- Given the URL template and the image description, the wrapper service accesses the image for specified day and spatial subset by HTTP Get SOAP protocol:
- Wrapper classes are available for geo-spatial data (incl. satellite), images, SQL servers, text files,etc.

ITR ReaSON

Information Landscape: Info System

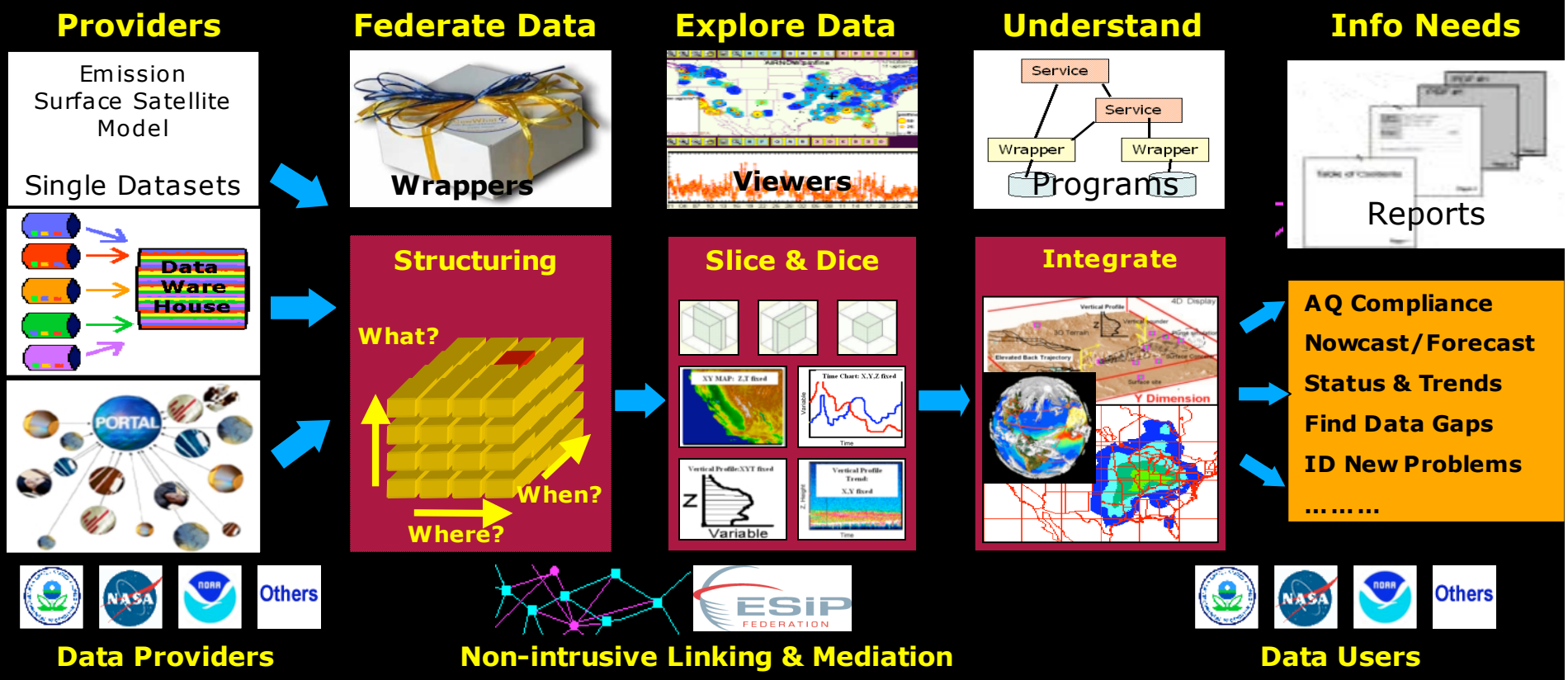
Data Access, Processing and Products



- The info system transforms the data into info products for each user
- In the first stage the heterogeneous data are prepared for uniform access
- The second stage performs filtering, aggregation, fusion and other operations
- The third stage prepares and delivers the needed info products

ITR ReaSON

Integrated Data System for Air Quality-IDAQ



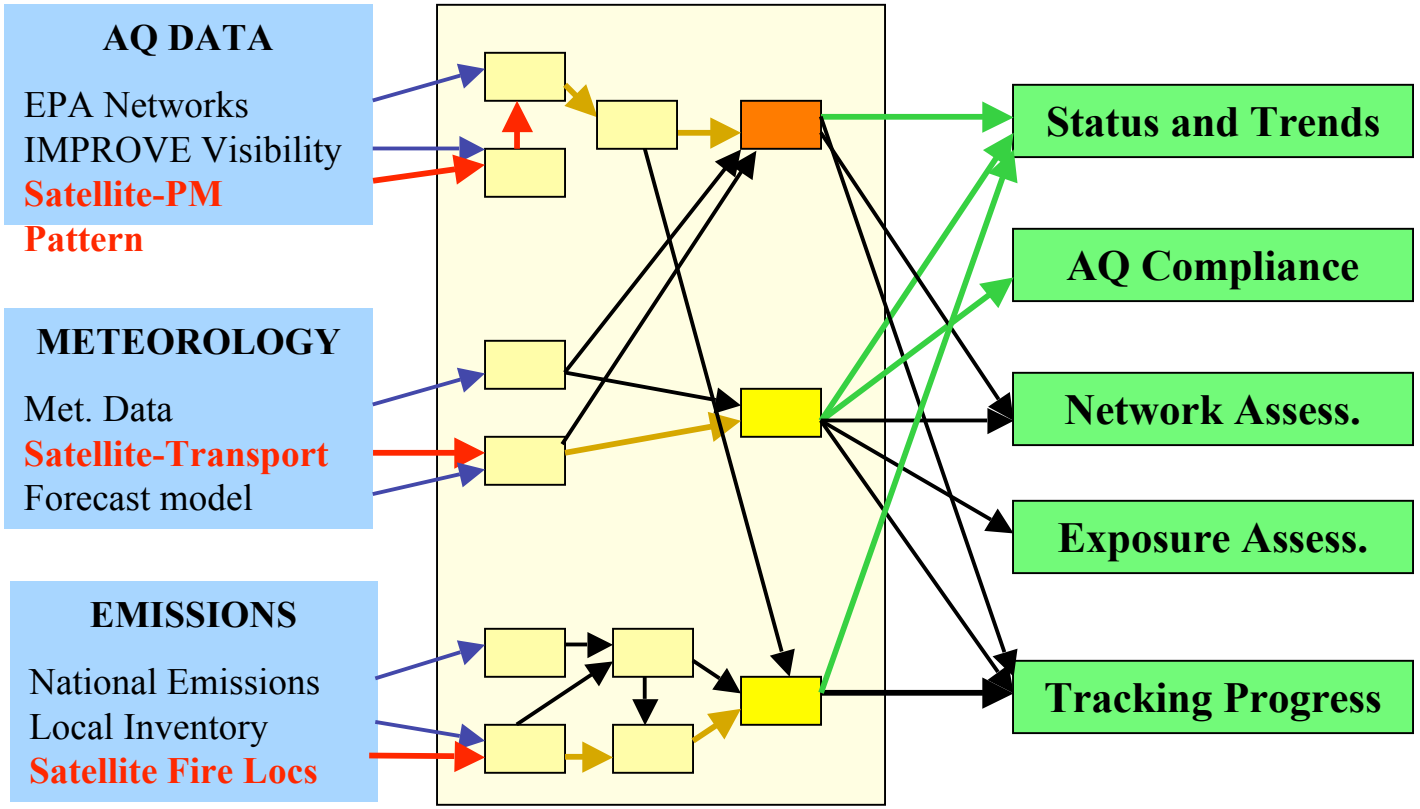
- The info system infrastructure needs to facilitate the creation of info products
- Providers supply the 'raw material' (data and models) for 'refined' info products
- The challenge is to design a general supportive infrastructure
- Simply connecting the relevant provides and users for each info product is messy
- Structuring the heterogeneous data into where-when-what 'cubes' simplifies the mess
- The 'cubed' data can be accessed and explored by slicing-dicing tools
- More elaborate data integration and fusion can be done by web service chaining
- This infrastructure support for IDAQ can be provided by the ESIP Federation

Data Flow & Processing in AQ Management

Primary Data
Diverse Providers

Data 'Refining' Processes
Filtering, Aggregation, Fusion

AQ Management Reports
'Knowledge' Derived from Data



ITR ReaSON

DataFed Description

DataFed Goals

Facilitate the **access and flow** of atmospheric data from provider to users
Support the development of **user-driven processing** value chains
Participate in specific application projects

Approach: Mediation Between Users and Data Providers

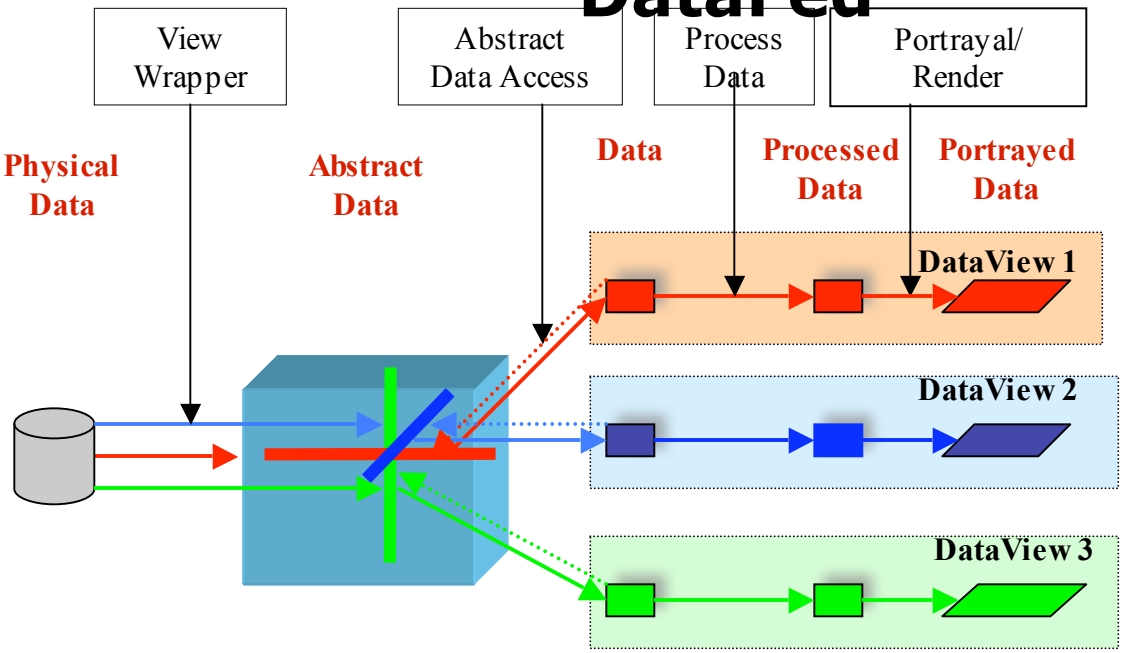
DataFed assumes spontaneous, **autonomous** emergence of AQ data (*a la* Internet)
Non-intrusively ***wraps*** datasets for access by web services
WS-based ***mediators*** provide homogeneous data **views** e.g. geo-spatial, time...
End-user programming of data access and processing through WS composition (limited)

Applications

Building **browsers and analysis tools** for distributed monitoring data
Serve as **data gateway** for user programs; web pages, GIS, science tools
DataFed is currently focused on the mediation of **air quality data**

ITR ReaSON

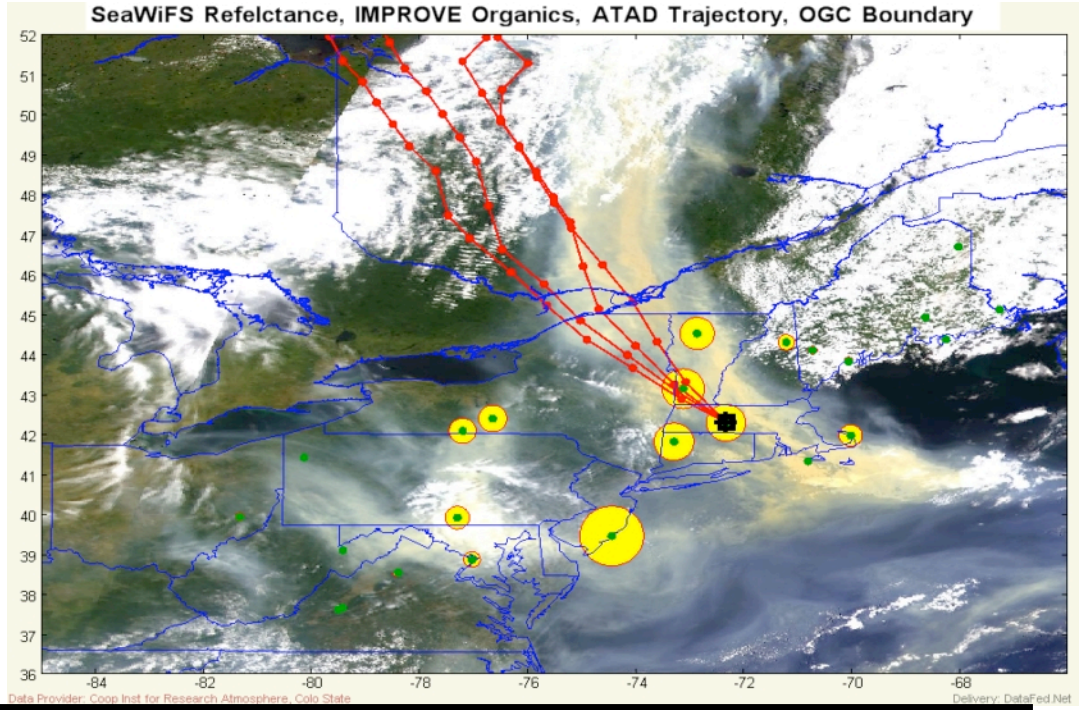
Generic Data Flow and Processing in DataFed



Physical Data	Abstract Data	Processed Data	View Data
*Resides in autonomous servers; accessed by view-specific **wrappers** which yield abstract data 'slices'*	*Abstract data slices are requested by viewers; uniform data are delivered by **wrapper services***	*Data passed through filtering, aggregation, fusion and other **web services***	*Processed data are delivered to the user as multi-layer views by portrayal and overlay **web services***

ITR ReaSON

Composition of Data Views from Distributed Data and Web Services



LAYERS

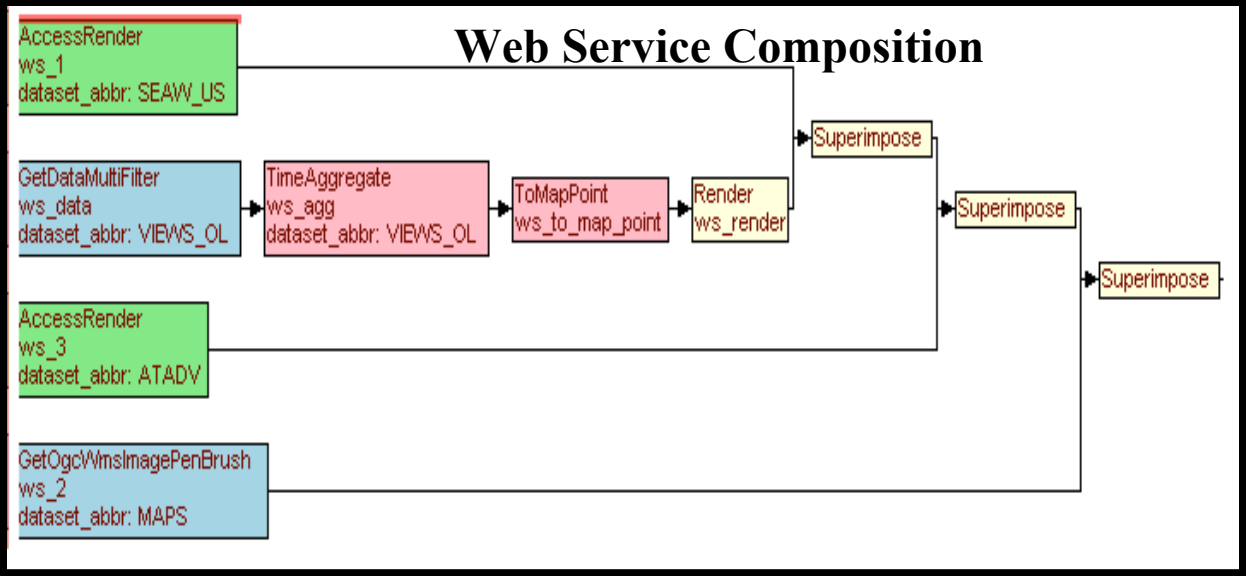
NASA SeaWiFS Satellite

RPO VIEWS Chemistry

NOAA ATAD Trajectory

OGC Map Boundary

Data Access | **Data Processing** | **Layer Overlay**



Service Flow Program for a VIEW

```
<ServiceFlow>
    <View>
      <MapImageMargin ref="ws_margins" />
      <AnnotateImage ref="ws_title" icon_text="T" label="Title" execute="true" />
      <StockAnnotation ref="DATAFED" />
      <StockAnnotation ref="CIRA" />
      <AnnotateImage ref="ws_cursor_annotation" icon_text="C" label="Cursor" execute="true" />
    </View>
    <Layers current="Traj_Grid" order="Traj_Grid Traj_Point Traj_Line MapImage_1">
      <Layer id="Traj_Line" dataset_abbr="VIEWS_OL" visible="false" visibility="1" click_target="true">
        <DataMapTimePoint ref="ws_data" label="Query Filter" />
        <DataMapTrajectory ref="ws_traj_render" show_data="false" icon_text="TR" />
        <AnnotateImage ref="ws_network_legend" icon_text="N" label="Network Legend" execute="true" execute_always="true" />
      </Layer>
      <Layer id="Traj_Point" dataset_abbr="VIEWS_OL" visible="true" visibility="1" click_target="true">
        <DataMapPoint ref="ws_point_data" show_data="false" />
        <DataMapPoint ref="ws_store_cursor_value" />
        <RenderMapPoint ref="ws_point_render" icon_text="R" label="point rendering" />
        <MapParamLocationAccessRender ref="ws_loc" execute_always="true" />
        <RenderLegend ref="ws_pntleg" icon_text="L" label="Point Legend" execute="true" execute_always="true" />
      </Layer>
      <Layer id="Traj_Grid" dataset_abbr="VIEWS_OL" visible="true" visibility="1">
        <DataMapTimePoint ref="ws_ref_data" label="Reference Query Filter" />
        <DataMapTrajectory ref="ws_ref_traj_grid" icon_text="TG" />
        <DataMapTimePoint ref="ws_data" label="Query Filter" />
        <DataMapTrajectory ref="ws_traj_grid" icon_text="TG" />
        <MapGridOperator ref="ws_mgo" />
        <RenderMapGrid ref="ws_rg" />
        <RenderLegend ref="ws_rainbow" icon_text="L" label="Rainbow Legend" execute="true" execute_always="true" />
      </Layer>
      <Layer id="MapImage_1" dataset_abbr="MapImage" visible="true" visibility="1">
        <MapImageAccessRender ref="ws_map_image" />
      </Layer>
    </Layers>
  </ServiceFlow>
```

View

Layer

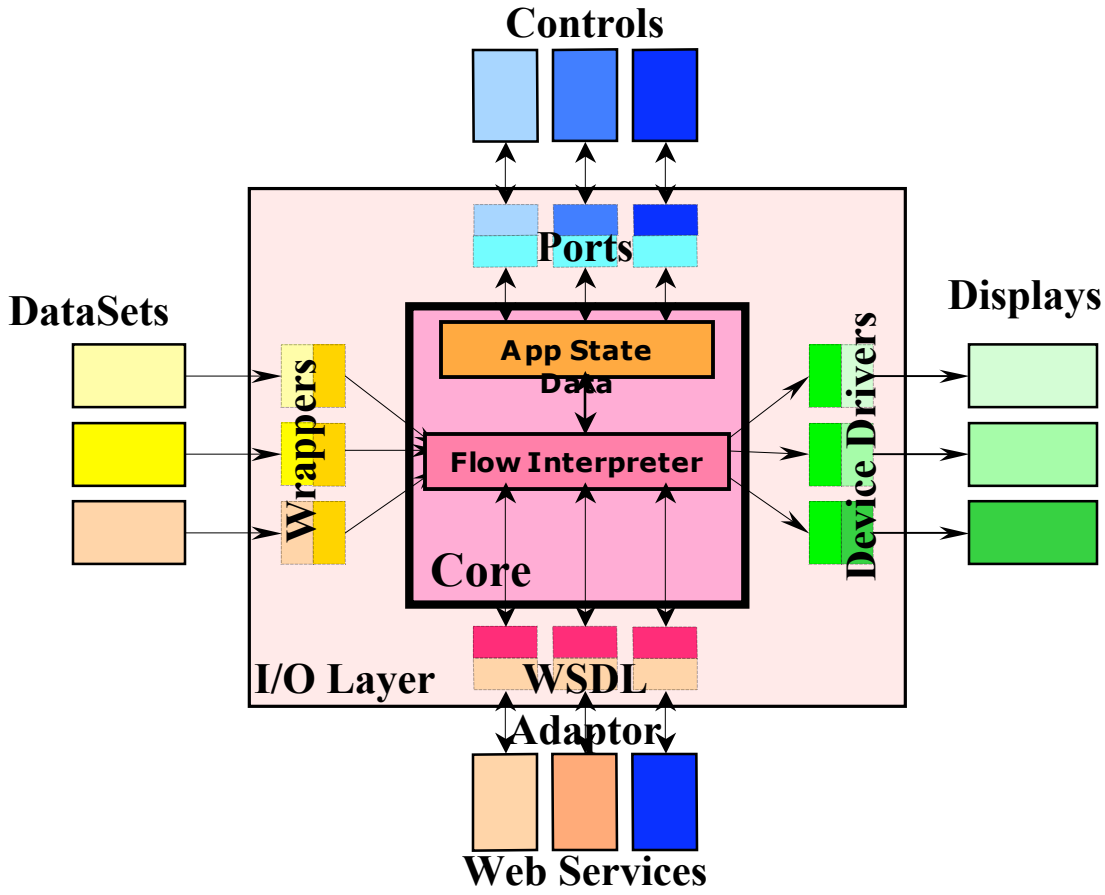
Layer

Layer

Layer

ITR ReaSON

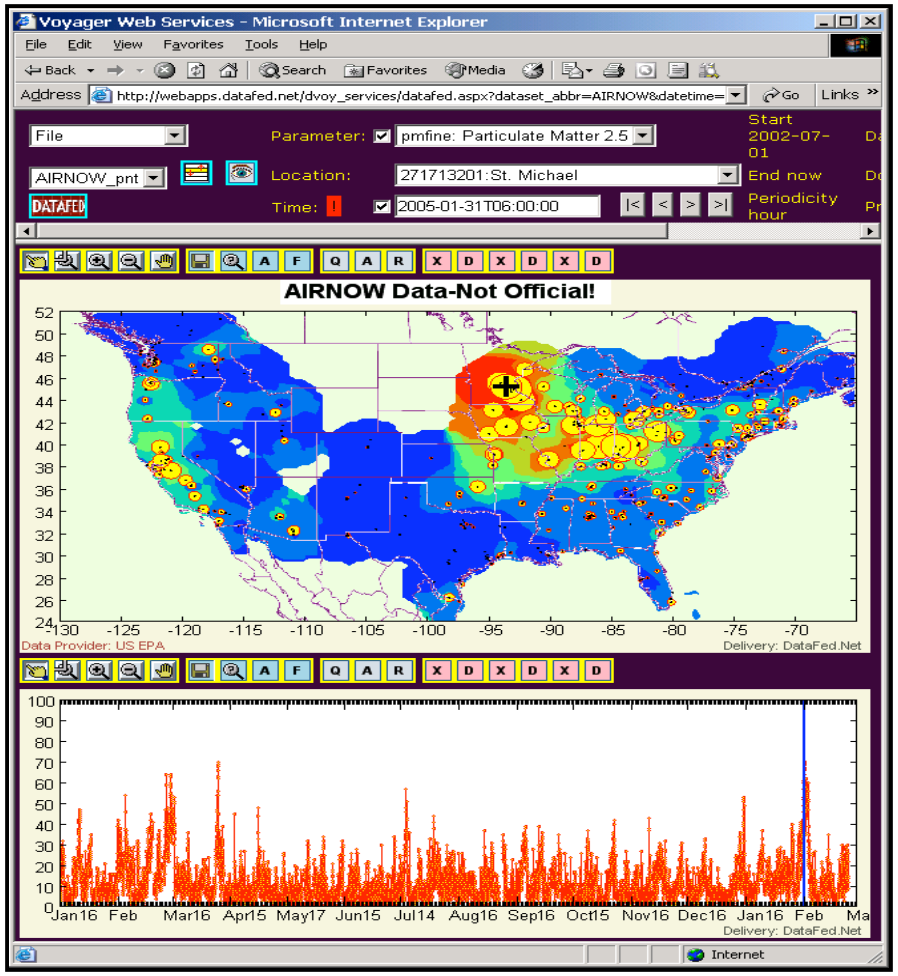
WS-Based Application Program Design

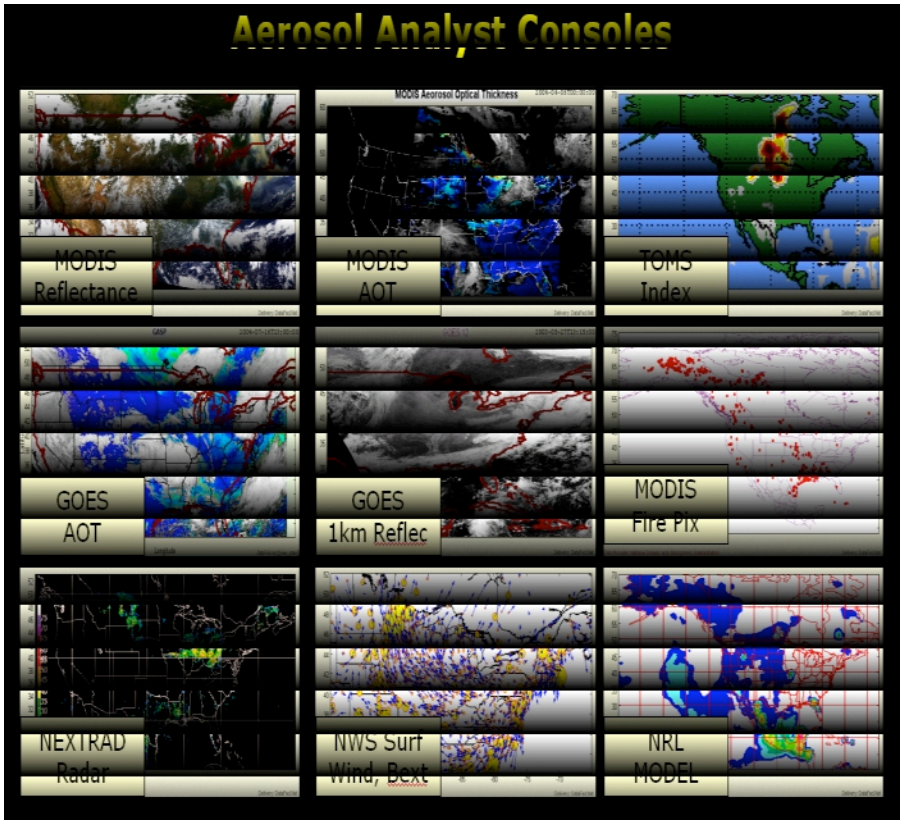


- The web-program consists of a stable **core** and adoptive **input/output** layers
- The core maintains the state and executes the data selection, access and render services
- The adoptive, abstract I/O layers connects the core to evolving web data, flexible displays and to the a configurable user interface:

ITR ReaSON

DataFed Tools for Episode Analysis





Viewer: General purpose spatio-temporal data browser and view editor applicable for all DataFed datasets

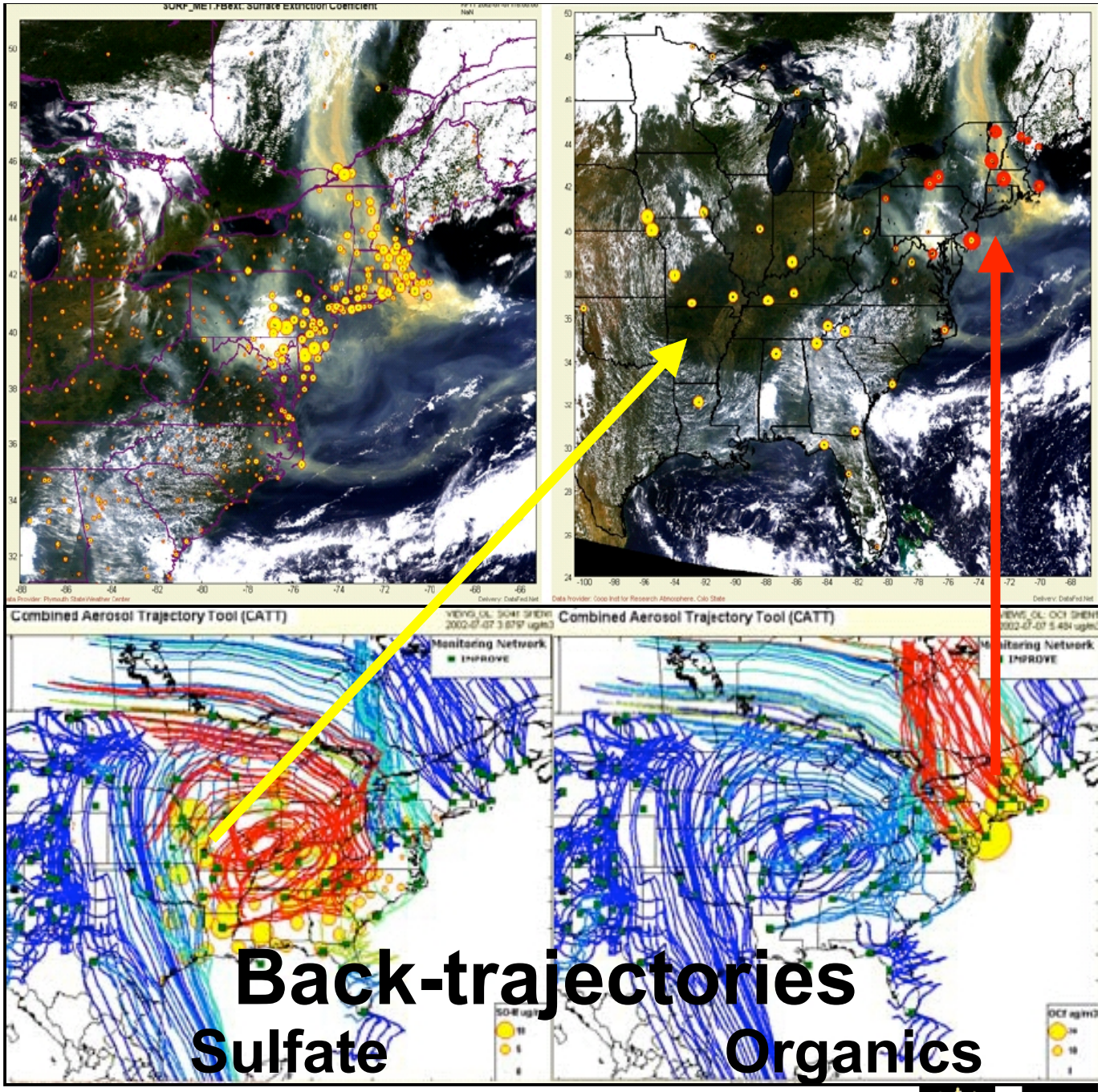
Consoles: Data from diverse sources are displayed to create a rich context for exploration and analysis

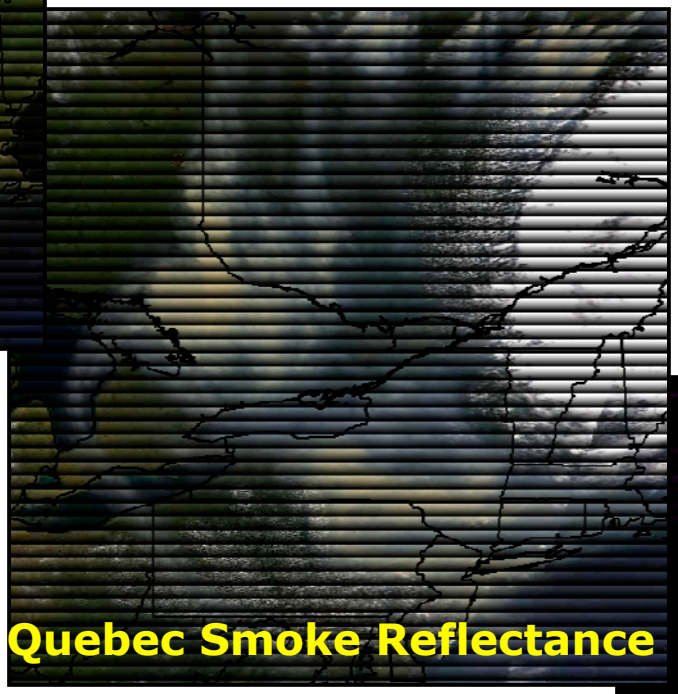
- Data Catalog
- Data Browser
- PlumeSim, Animator
- Combined Aerosol Trajectory Tool (CATT)

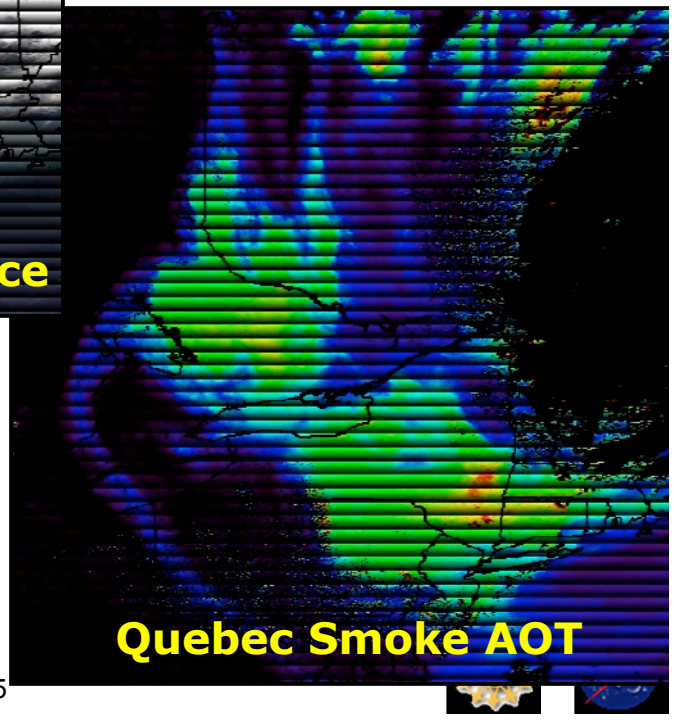
ITR ReaSON

CATT

Combined Aerosol Trajectory Tool







Air and Waste Management Association Professional Development Course

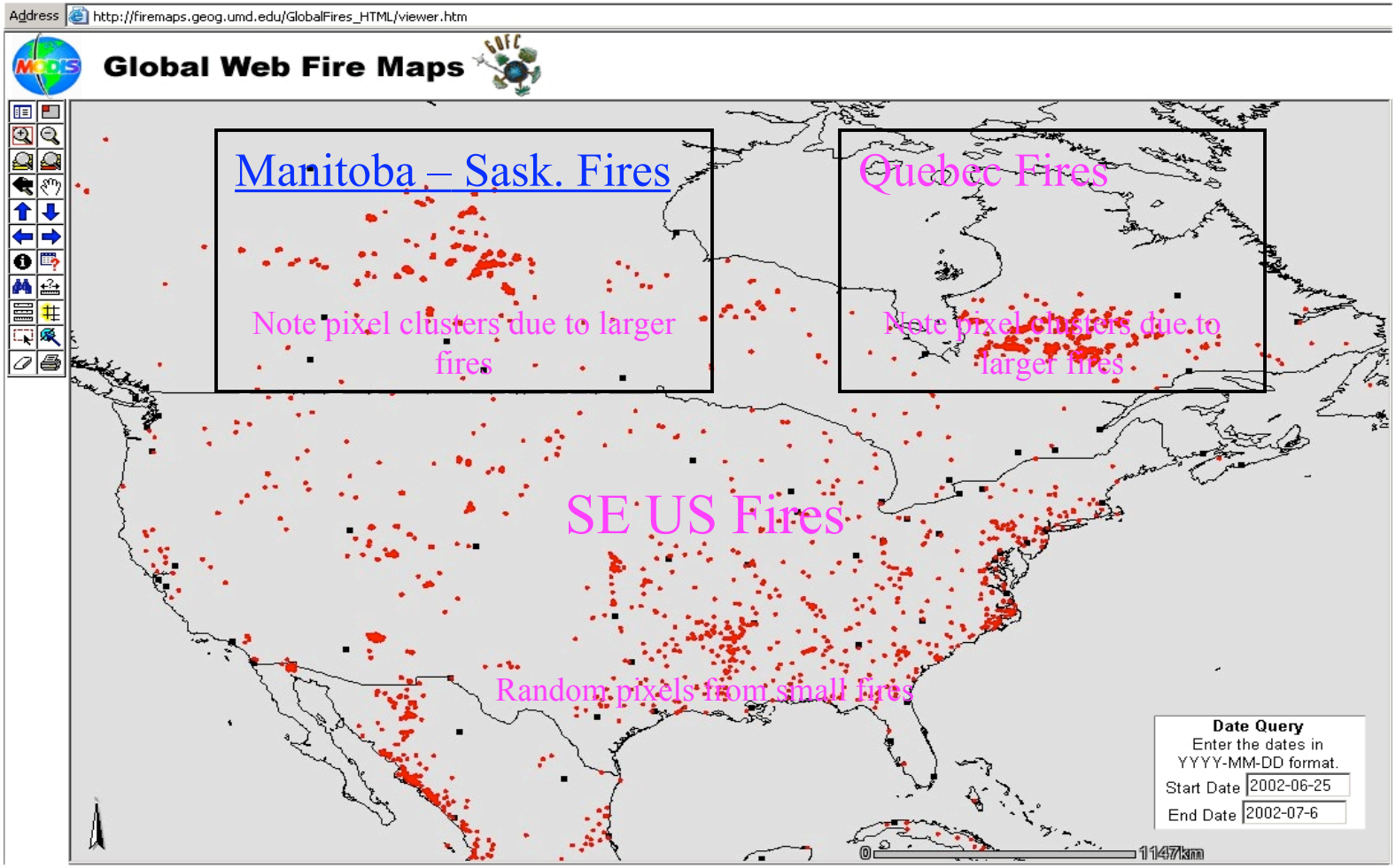
AIR-257: Satellite Detection of Aerosols

Instructor:

Rudolf Husar, Ph.D. Professor of Mechanical Engineering
Washington University, St. Louis, MO

October 25, 2004, 9:00 a.m. - 12:00 p.m. Asheville, NC

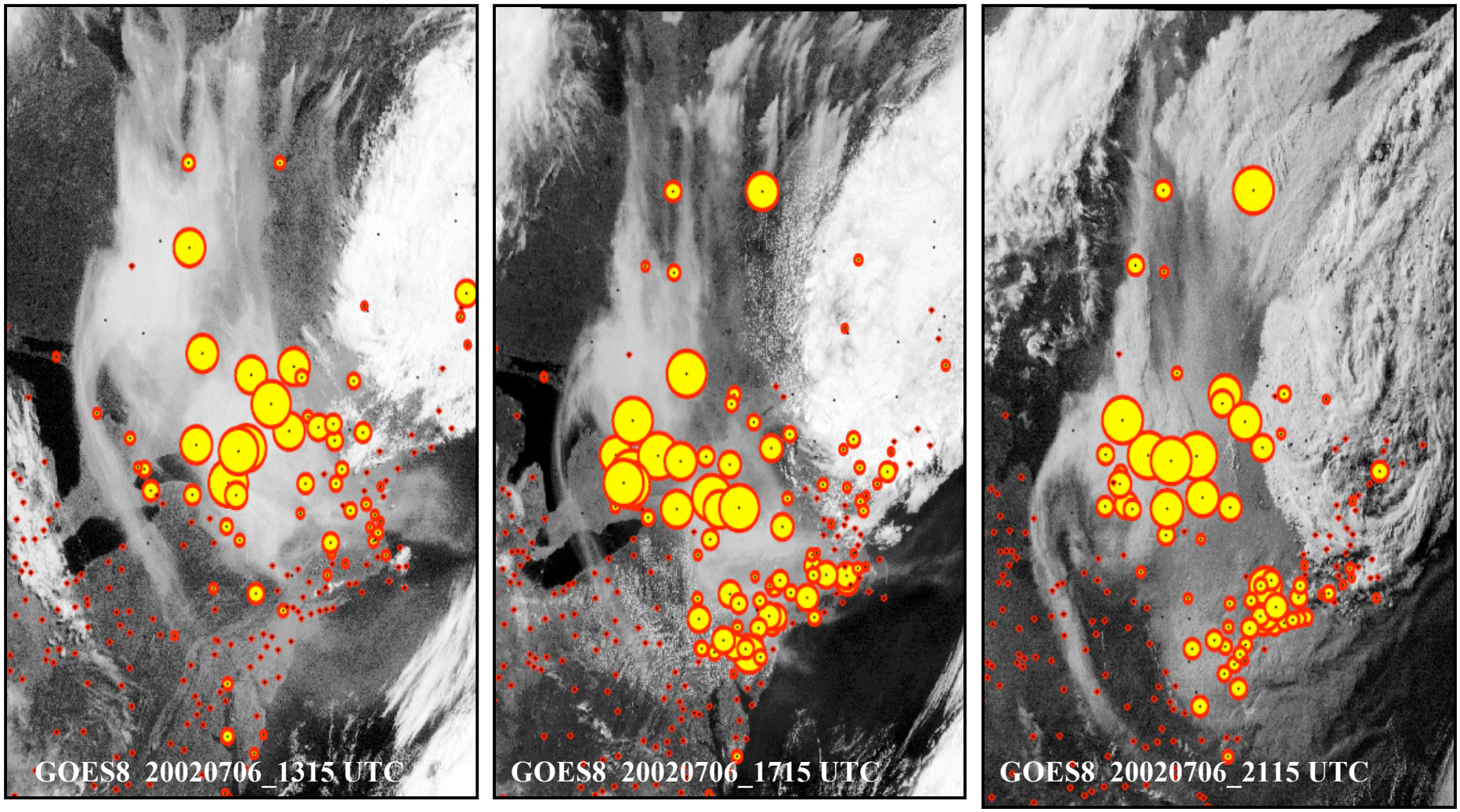
Fire Pixels from MODIS (UMD), June 25-July 6, 2002



- Several satellite sensor (MODIS, GOES, AVHRR, ATSR…..) detect the location of most fires - DAILY
- These 'fire pixels' can be used as sensor-based inputs to regional/global models, e.g. NAAPS
- However, the quantity of smoke emitted from the from the 'fire pixels' can not be estimated well
- Hence, real-time model simulation of smoke transport is limited by the smoke emission estimates

ITR ReaSON

GOES 8 – METAR



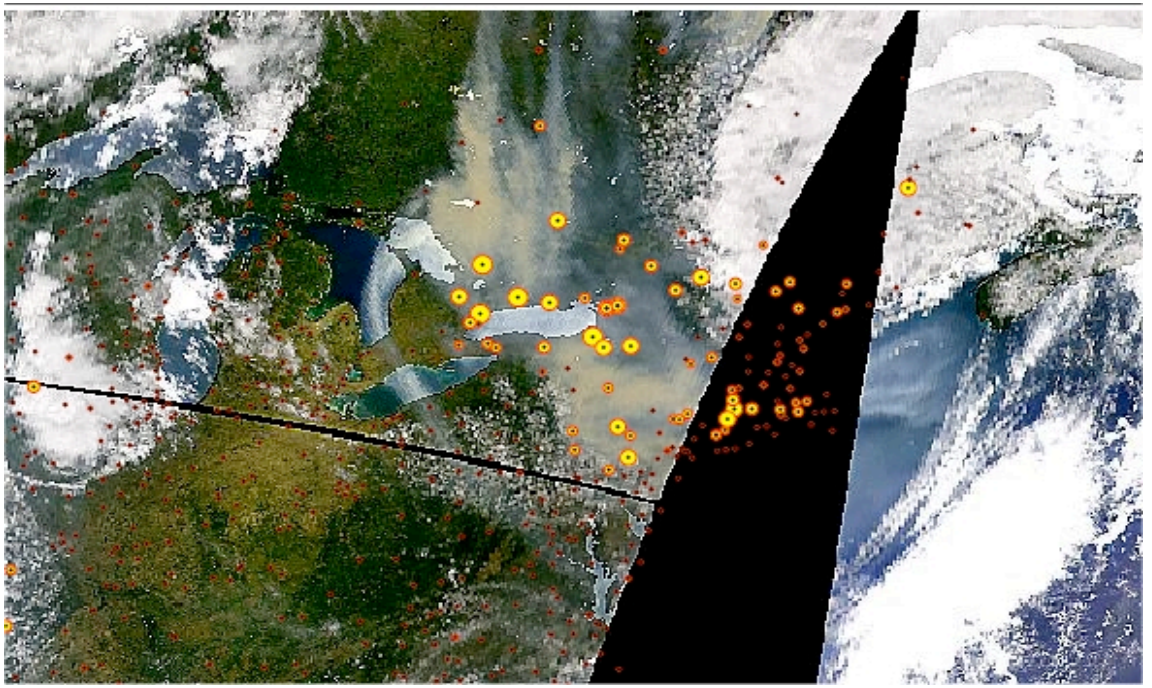
July 6, 2002 8:15, 12:15, 16:15 EST

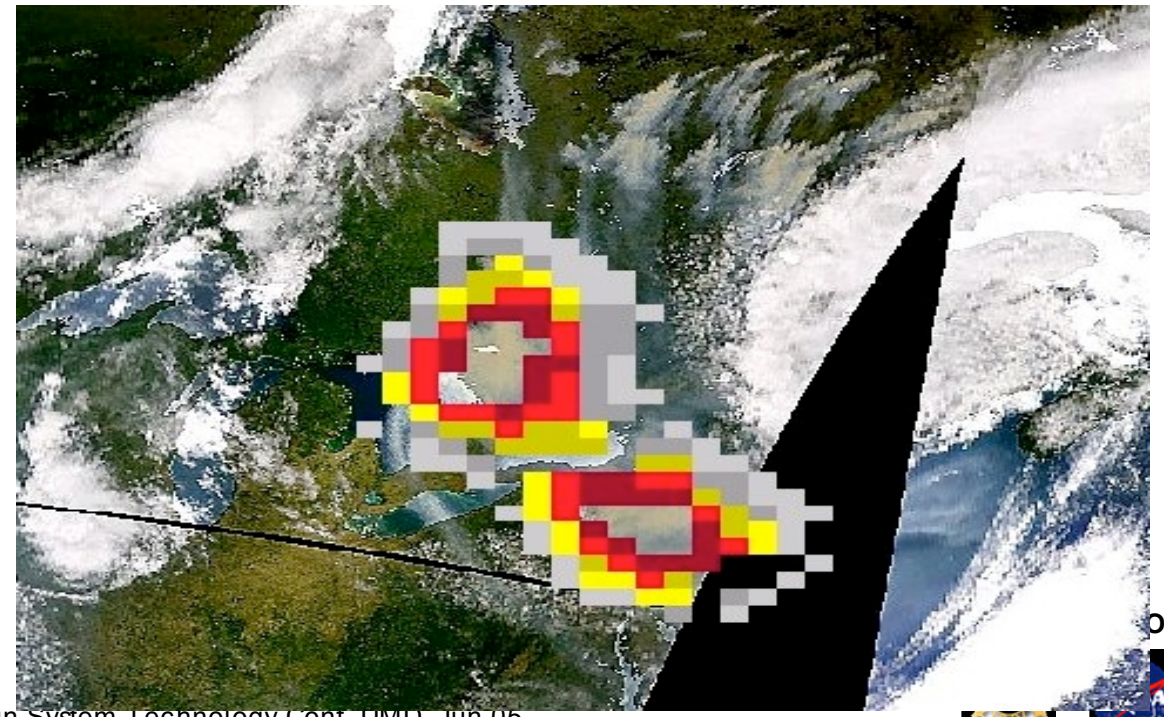
ITR ReaSON

SeaWiFS & ASOS & TOMS

SeaWiFS & TOMS

- SeaWiFS & ASOS Yellow circles proportional to ASOS Bext





- SeaWiFS & TOMS absorbing aerosol index.
- Notes: Yellow color (absorbing in blue?); no TOMS for fresh smoke

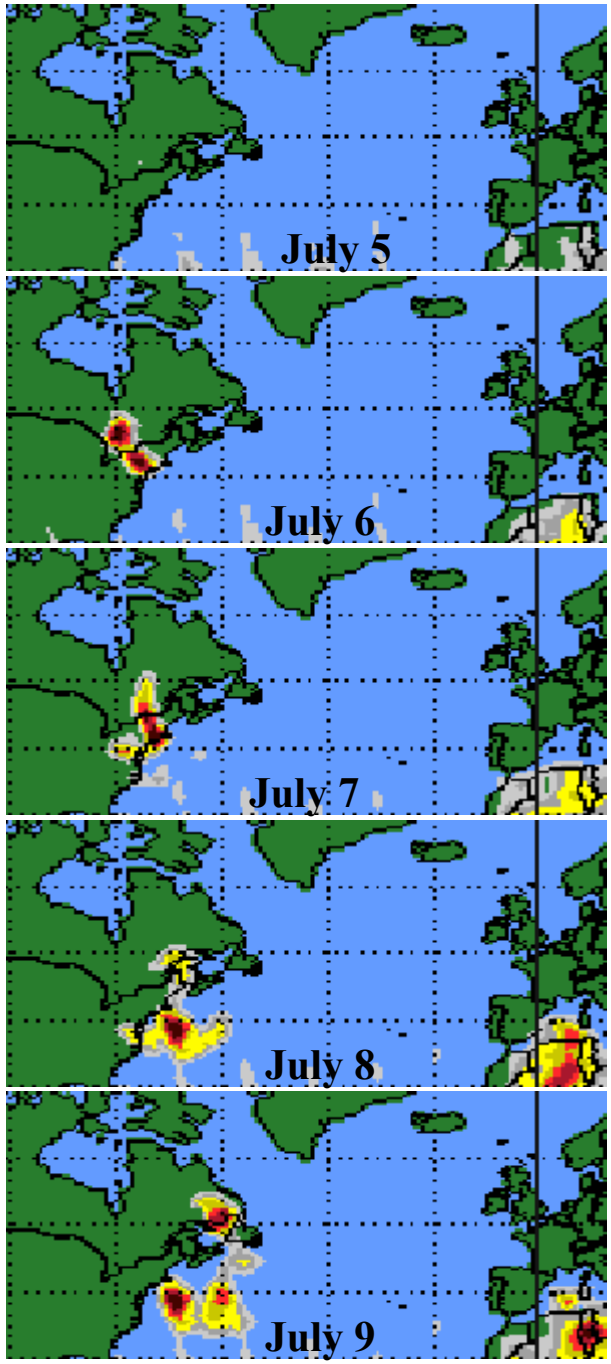
TOMS: The Big Picture

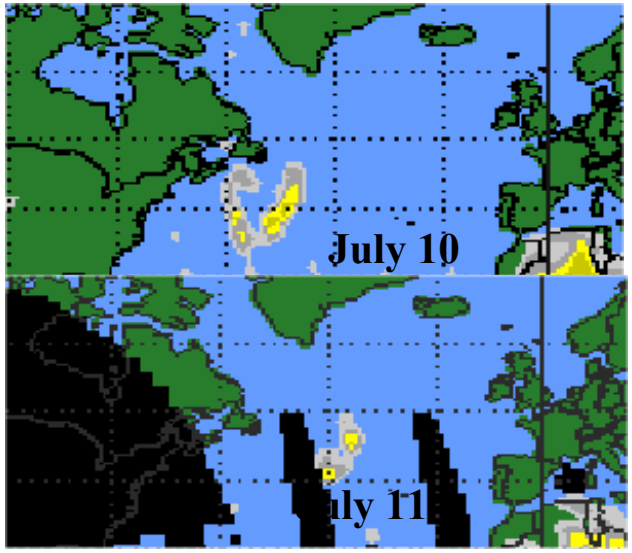
Absorbing Aerosol Index

July 5: The near-source, low level smoke is not detected by TOMS

July 6-7: Smoke plume signal is very intense over S. Ontario and NE US.

July 8-9: Transport to the Atlantic.
Where will the smoke reach Europe?
How intense, will it be detectable?
Would anyone run Hysplit, ATAD?

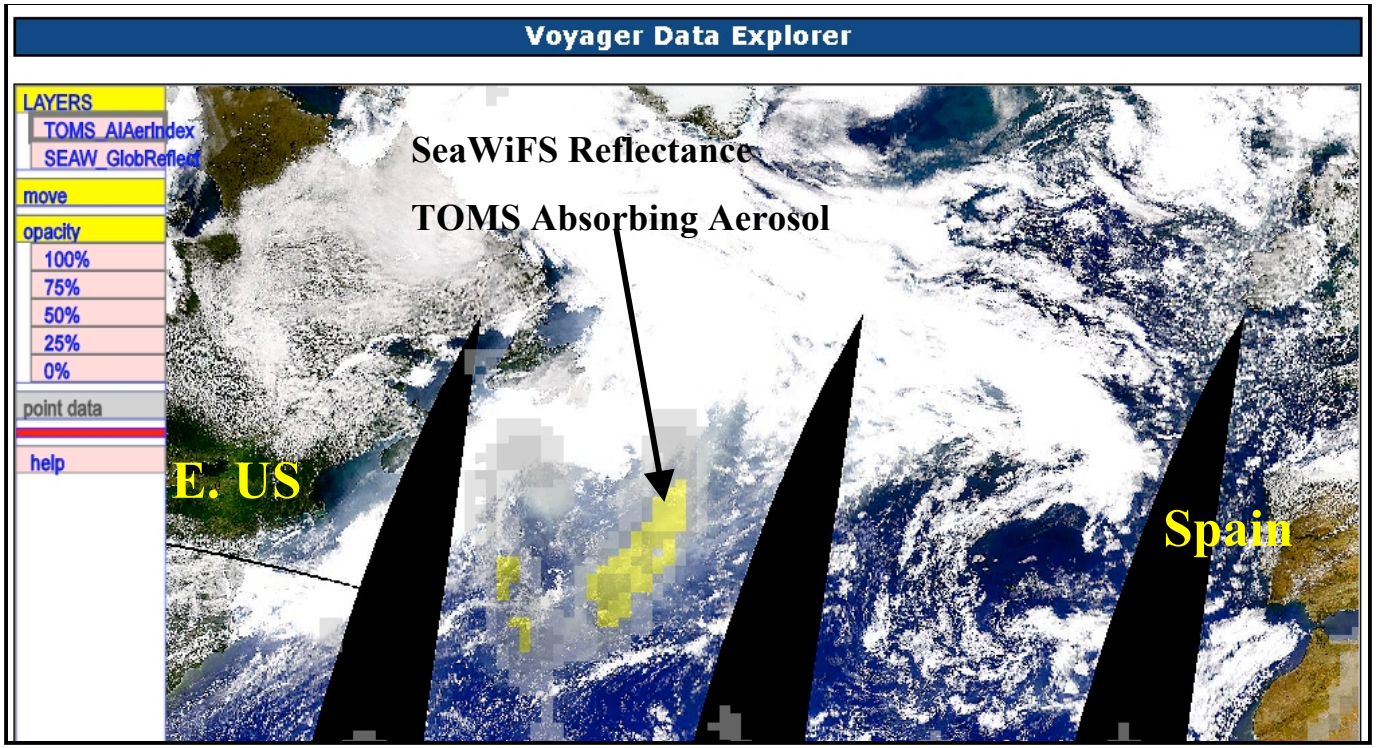




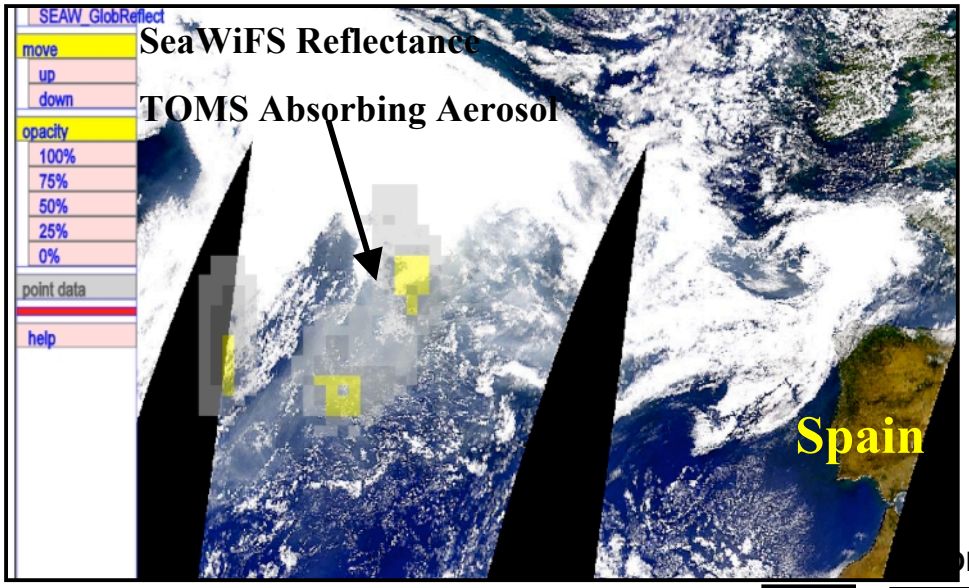
ITR ReaSON

Trans-Atlantic Transport of Quebec Smoke

July 10: Quebec smoke over Mid-Atlantic



July 11: Smoke approaching Europe



Summary

Web Services in Practice:

- Access to distributed, heterogeneous data
- Process chaining is practical
- We have replaced the data analysis tools with WS versions
- However, we are still not interoperable with other WS -shops

The Future: Loosely Coupled Info Systems

- Seamless Data Access:
  - **Data 'Wrappers'**
  - Standard protocols (OGC, OpenDAP…)

- Seamless Service Linking:
  - **Services Adapters**
  - Standard protocols

ITR ReaSON

The Future:
Value-Added Processing in Service Oriented Architecture

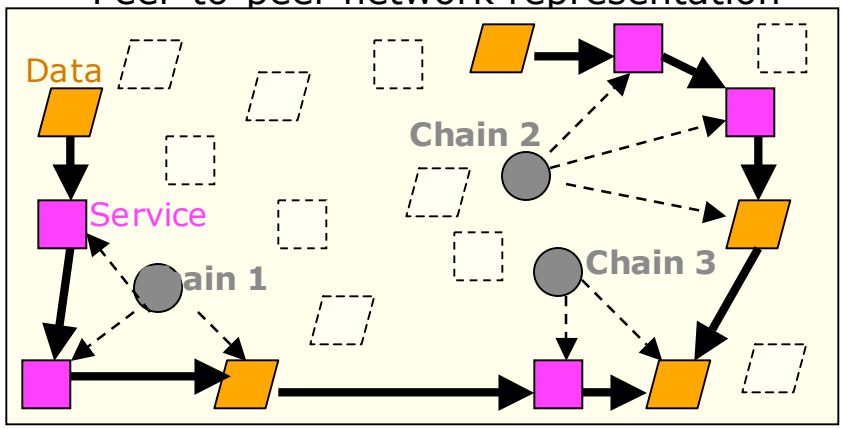
Data, services and users are distributed throughout the network

Users compose data processing chains form reusable services

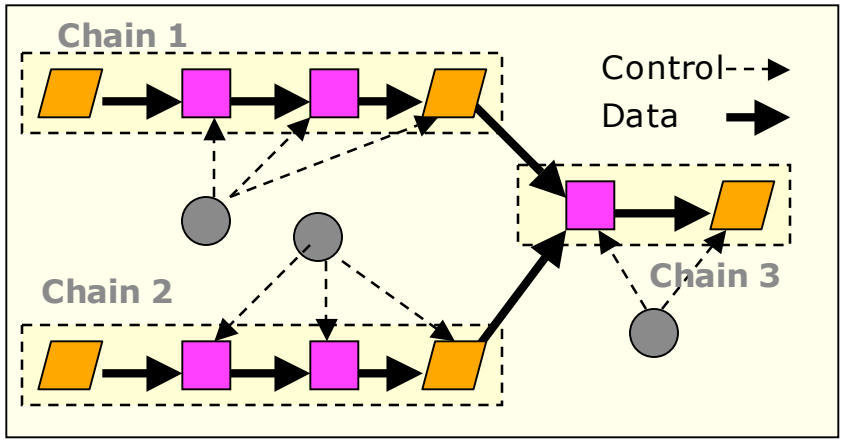
Intermediate data are also exposed for possible further use

Chains can be linked to form compound value-adding processes

Peer-to-peer network representation



Service chain representation



User Carries less Burden

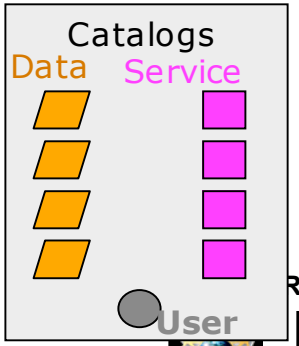
In service-oriented peer-to peer architecture, the user is aided by software 'agents'

User Tasks:

Find data and services

Compose service chains

Expose output



ReaSON